Introduction to Survival Analysis

Emiliano Rossi

Cardiologist. Department of Investigation. Hospital Italiano de Buenos Aires. City of Buenos Aires, Argentina.

Acta Gastroenterol Latinoam 2022;52(1):11-14

Received: 19/01/2022 / Accepted: 18/03/2022 / Published online: 30/03/2022 / https://doi.org/10.52787/agl.v52i1.164

What is Survival Analysis?

In many clinical situations, we are interested not only in studying whether an event occurs, but also in determining at what point it take place during the follow-up.

Survival analysis is a set of statistical procedures for data analysis in which the outcome variable is the time until an event occurs.

What Is It for?

The first applications of this analysis studied the time from the start of treatment to death, hence the name "survival". Subsequently, it was extended to other situations of clinical interest (time to readmission, treatment abandonment, return to work after surgery, etc.), so that a more appropriate name is time-to-event analysis.

The objectives of survival analysis are: a) to estimate and interpret survival, b) to compare survival between different groups, and c) to evaluate the relationship of different explanatory variables with survival.[1]

What Is the Endpoint Evaluated?

Usually a single event of interest is studied, although a combined endpoint may be established (the first component to be considered for the analysis). In other more complex situations, recurrent or competitive events are considered.[1]

In What Study Designs Is Applied?

The designs where the survival analysis methodology can be applied are cohort studies and clinical trials.

What Information Needs to Be Recorded for Its Implementation?

To apply this analysis, it is necessary to determine at the beginning of the study which patients are susceptible to the event, excluding those who are not. Then record the date on which the event of interest occurs or the end date of follow-up, in the cases where the event did not occur.

What Is the Survival Function and the Hazard Function?

They are two closely related concepts. The survival function *S (t)* is defined as the probability of surviving at a given time. The hazard function h *(T)* is the conditional probability of the event occurring at a certain time, given that the individual has survived to that time.[2] The graph of the survival function is the survival curve.

Correspondence: *Emiliano Rossi*
Email: emiliano.rossi@hospitalitaliano.org.ar

What Is Censorship?

It is a relevant concept to consider when performing a survival analysis. The time to the occurrence of the event is not visible in all patients, it is censored because follow-up has been interrupted (i.e., the patient has been lost to follow-up). There is no way of knowing when the event will occur thereafter. The situations in which survival time can be censored are: when there is loss to follow-up, when there is withdrawal from the study, and when the study ends and the patient has not yet presented the event.[1]

What Is the Advantage of Applying this Analysis?

The advantage of using the survival analysis is that it allows us to consider the information available on all patients, not just those who reached the end of the follow-up. Each patient provides information during the period they were in the study, even if they later lost follow-up. Thus, patients who present early censoring provide less information than those who are followed for a long time. Therefore, all observations contribute some information. On the other hand, if only patients who completed the study were included in the analysis, a bias would be introduced.

What Are the Statistical Methods Used?

The best-known methods that allow handling censorship in survival analysis are the Kaplan-Meier curves and the Cox proportional hazards model.

Kaplan-Meier curves plot the proportion of patients who have survived over time in each treatment (exposure) group. The height of the Kaplan-Meier curve at the end of each time interval is determined, by taking the proportion of patients who were event-free at the end of the previous interval and multiplying it by the proportion of patients who survived at the end of the current interval. This iterative process of multiplying probabilities begins at the first-time interval and continues to the last. The survival curve is not modified when an observation is censored, but, in the following period, those who have been censored are excluded from the number of people at risk.

Comparison of the Kaplan-Meier curves is carried out using the log-rank test.[2] This test examines the null hypothesis that there is no difference between the survival curves of the different groups, considering the observed events for each group.

The Cox proportional hazards model allows solving the censoring problem and, in addition, adjusting for covariates (e.g., confounders). The summary measure of the model is the hazard ratio (HR). The HR is the ratio between the hazard (risk) of the group of interest and the reference group. It allows us to estimate how much the risk of suffering the event increases for each unit increase in the explanatory variable. For example, if the HR is 1.25 for age (in years), this means that for each increase in one year the individual has a 25% higher risk of presenting the event of interest.

What Are the Limitations of the Cox Proportional Hazards Model?

This model requires two assumptions to be met.[3] The first is that the censored data must be independent of the outcome of interest. This is not true in the case of patients who are lost to follow-up, because of the event of interest. The second assumption is that the hazards are proportional, meaning that the HR is constant throughout the duration of the study. This means that, if the risk increases (or decreases) in the exposed group, it should concomitantly increase (or decrease) in the comparator group, keeping the relationship between the two-remaining constant. If these assumptions are not met, the model results are not valid.

When Is the Hazard Proportionality Assumption not Met?

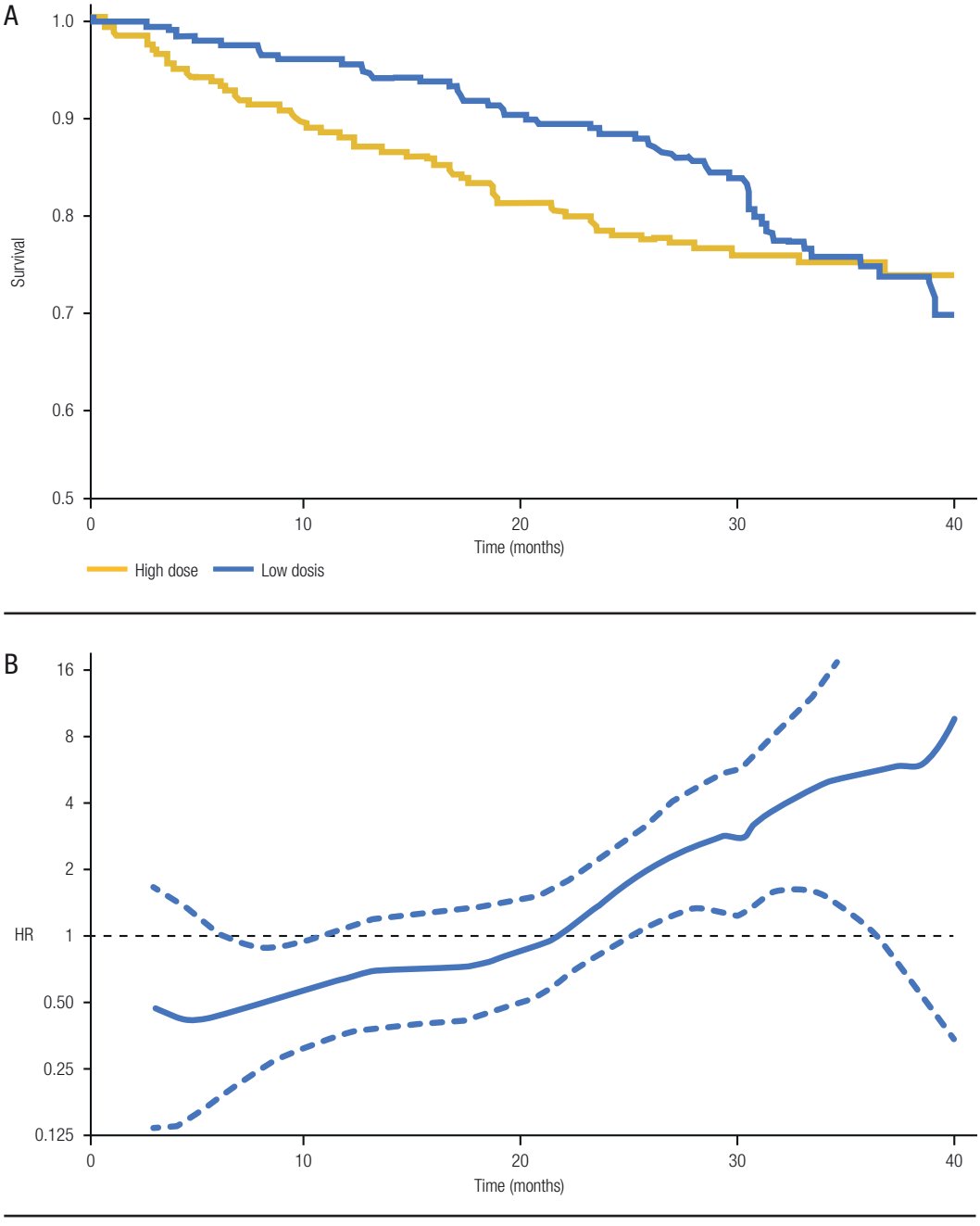
The HR is not constant when the effects of the treatment change over time or when the susceptibility to suffer the event varies among individuals in one of the groups (giving rise to the phenomenon of exhaustion of susceptibles).[4] These situations are frequent in clinical studies. It must be taken into account that, even if tests are applied to evaluate the proportionality of the hazards, they may not have sufficient power.[4, 5]

For practical purposes, we should suspect that the hazard proportionality assumption is not fulfilled when the survival curves show obvious changes in their slopes. Figure 1A shows the survival of two groups of patients with multiple myeloma treated with low and high doses of dexamethasone (ECOG Study): in the initial period, the survival curves are divergent, then they tend to be parallel and finally, converge (even intersecting). If we observe the HR graph throughout the follow-up (Figure 1B), we notice that the HR is not constant. Initially, it is less than 1 (indicating that the hazard of the low-dose group is lower), then towards the middle of follow-up it is equal to 1 and finally, it is greater than 1. The estimated HR in this study was 0.87, which should not be interpreted as a uniform 13% reduction in mortality for the low-dose group.[5]

What to Do If the Hazard Proportionality Assumption Is Violated?

In these cases, we should not use HR but rather other measures to compare the differences in survival between groups. One of the alternatives available is the restricted median survival time ratio. The restricted median survival time (RMST) represents the group average of event-free time during the period. It is calculated by estimating the area under the Kaplan-Meier curve up to a given time. In the aforementioned study (ECOG), the RMST at 40 months of the low-dose group has an area under the curve equal to 35.1 months (meaning that a patient in this group is expected to live 35.1 months out of 40 months of follow-up). In contrast, in the high-dose group, the area under the curve represents 33.3 months. Thus, the ratio of RMST = 35.1/40 = 1.06 (95% CI 1.00-1.13).[5]

Figure 1. A) *Survival curves according to treatment,* ***B)*** *HR estimate (and 95% CI) during follow-up.*



Cite this article as: *Rossi E. Introduction to Survival Analysis. Acta Gastroenterol Latinoam. 2022;52(1):11-14. https://doi.org/10.52787/agl.v52i1.164*

References

1. Kleinbaum DG, Klein M. Survival Analysis: A Self-Learning Text. New York: Springer; 2012.
2. Bewick V, Cheek L, Ball J. Statistics review 12: Survival analysis. Critical Care 2004;8:389-395.
3. Tolles J, Lewis RJ. Time-to-Event Analysis. JAMA 2016; 315(10): 1046-1047.
4. Stensrud MJ, Hernán MA. Why Test for Proportional Hazards?. JAMA 2020;323(14):1401-1402.
5. Uno H, Claggett B, Tian L, *et al.* Moving Beyond the Hazard Ratio in Quantifying the Between-Group Difference in Survival Analysis. J Clin Oncol 2014;32:2380-2385.